Combined Approach of Group-Contribution and Molecular Modeling for the Estimation of Pure Component Properties

J. Marrero and R. Gani[C,S]

CAPEC, Department of Chemical Engineering, Technical University of Denmark
DK-2800 Lyngby, Denmark

rag@kt.dtu.dk

A combined approach of group-contribution (GC) method and molecular modeling (MM) is proposed for the estimation of important properties of pure organic compounds for a wide range of organic compounds. The property values are estimated by a combination of a multilevel GC method and MM methods in where each type of method complements the other by providing the "missing" data or information. In this way, it is possible to start at the MM level and go to the GC level or vice versa.

The GC method allows the estimation of several important properties such as solubility parameter, water solubility, octanol/water partition coefficient, toxicity, melting point and enthalpy of fusion and others. Estimation is performed at three levels. The primary level uses contributions of simple groups that allow describing a wide variety of organic compounds but cannot represent adequately proximity effects or isomer differences in large and polyfunctional compounds. This limitation is overcome by using the higher levels of estimation, which involve groups that provide more information about molecular fragments whose description in primary level is insufficient. The group contribution parameters have been calculated from regression using a data set of about 9000 compounds ranging from C=3 to C=100.

Despite the considerable accuracy and range of applicability of the GC method, the estimation of some properties such as melting point or enthalpy of fusion needs an additional improvement in the cases of stereoisomers of very complex molecules (e.g. steroids and large aminoacids). Therefore, a further step involving molecular modeling is applied. As input, this step receives the atom-atom connectivity matrix of a given molecular structure, which is transferred to a molecular modeling (MM) program (e.g. MOPAC or Chem3D). MM methods are then used to check if extrapolations from the GC method with group parameters outside the range of their "know" applicability are qualitatively correct and to define new groups and their contributions when they are needed but do not exist in the GC method.